

Gene Expression Microarray Analysis of Archival FFPE Samples

Application Note

Gene Expression Analysis

Authors

Anne Bergstom Lucas
Gary Lin

Agilent Technologies
Santa Clara, CA USA

Abstract

Microarray-based gene expression hybridization is a powerful and proven technique for studying differential gene expression signatures in cancer. RNA extracted from freshly frozen tissues is optimal for microarray analysis; however, in many cases, formalin-fixed paraffin-embedded (FFPE) tissues are the only samples available. To work with these difficult samples, Agilent has developed an FFPE sample protocol optimized for use with Agilent gene expression microarrays. To test the effectiveness of this protocol, total RNA was extracted from a quadruplicate set of colon tumor (adenocarcinoma), and matched normal colon FFPE and fresh frozen tissues. RNA purity and quality were determined, and the RNA was amplified using a system in which a reverse transcription-generated cDNA library is subjected to limited PCR amplification. The resulting cDNA target was chemically labeled. Analysis of the microarray results revealed a reduced sensitivity of detection for the FFPE-derived RNA samples; however, the lists of detected genes showed high overlap between the FFPE and frozen samples. Comparison of gene expression signals demonstrated excellent assay reproducibility between the technical replicates. Principal component analysis of array signals demonstrated clear separation between the tumor and normal samples with grouping of replicate samples. Statistical analysis of the expression level variations between tumor and normal samples revealed that both the FFPE and frozen samples had more than 2200 genes with a fold-change greater than 2 with a corrected p-value of 0.05 or less. The FFPE and frozen samples showed good concordance of tumor vs. normal \log_2 ratios with few anti-correlated genes.

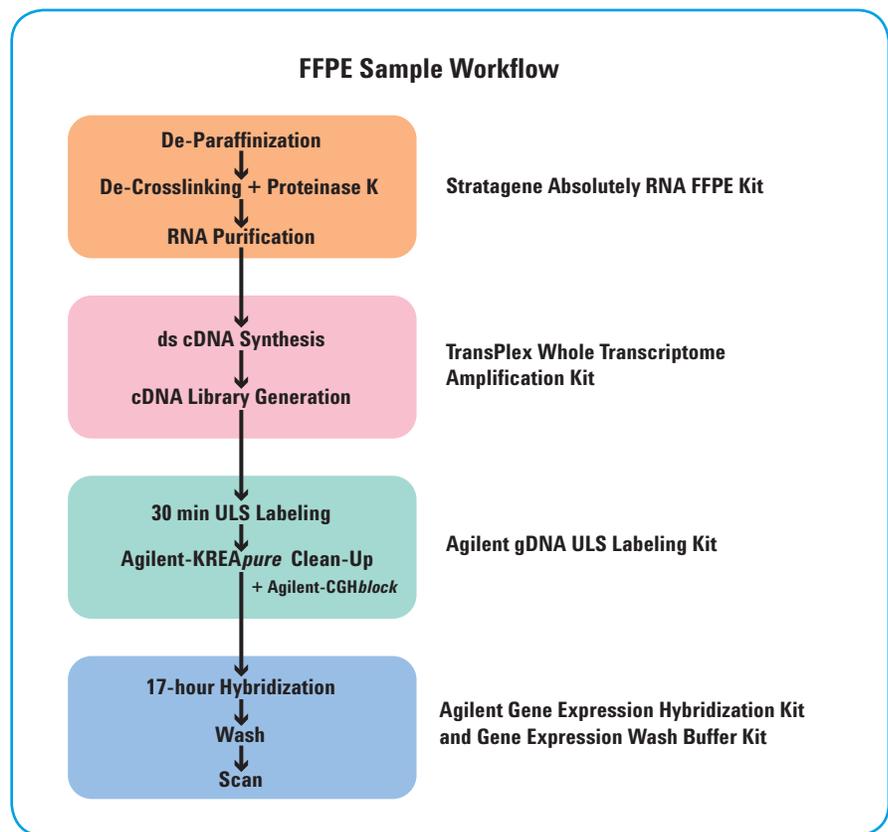


Agilent Technologies

Introduction

Microarray analysis is a widely used technology for studying gene expression on a global scale. Studies have shown that microarray gene expression profiling results in improved diagnosis and risk stratification in cancer.¹ Recently, a 70-gene expression profile was established as a powerful predictor of disease outcome in young breast cancer patients.² These studies utilized high quality RNA extracted from fresh frozen tumor tissue. Many laboratories are trying to develop methods to allow a similar degree of gene expression profiling using FFPE samples.³ FFPE samples represent the largest source of archival biological material available for large retrospective prognostic studies of human cancer, with over 400 million FFPE samples estimated in existence. Challenging samples such as degraded or FFPE tissues have remained difficult for gene expression microarray technology.

RNA isolated from FFPE samples is difficult to use in standard enzymatic procedures due to cross-linkage between nucleic acids and proteins, covalent RNA modification by the addition of mono-methylol groups to the bases, dimerization of adenine groups by methylene bridging, and RNA fragmentation and degradation.⁴ The extent of damage increases with the amount of time after exposure to the fixative and is also dependent on the particular fixing protocol employed.³ The situation is further complicated because there are currently no consistent standards for tissue fixation. The deleterious effects of chemical fixatives on nucleic acids result in impure RNA and low yields when standard RNA isolation methods are used.⁴



Box 1. FFPE sample workflow. Flowchart indicates the various kits and steps involved in the analysis of FFPE samples with the Agilent gene expression microarray workflow.

For this study, we focused on FFPE and fresh frozen samples from adenocarcinoma and normal adjacent tissues (NAT). These matched normal and malignant colon tissues enable comparison of gene expression signatures between the two types of tissue storage. RNA was extracted from each type of tissue using extraction kits especially designed for each storage condition. The total RNA was DNase-treated and checked for DNA contamination by qPCR, and amplified cDNA was generated with a whole transcriptome amplification system. The amplified cDNA was labeled using a non-enzymatic one-step labeling method called the Universal Linkage System (ULS). The ULS system was chosen to label the amplified cDNA because it is a direct labeling method and is independent of fragment length.

Dye-labeled target was applied to 4x44K Whole Human Genome microarrays, and the microarrays were scanned using an Agilent DNA Microarray Scanner equipped with extended dynamic range (XDR) software. Microarray data analyses were performed using the Agilent GeneSpring GX 10.0 software program. An overview of this workflow for Agilent gene expression microarray analysis of FFPE samples is presented in **Box 1**.

An important consideration regarding the value of gene expression data generated from compromised FFPE RNA is whether these results yield informative biological information. To address this question, we compared expression results obtained from FFPE-derived RNA to RNA collected from matched fresh frozen samples.

Materials and Methods

RNA isolation and analysis

Matched normal and malignant colon tissues, from both fresh frozen and FFPE samples, excised and stored in 2006, were obtained from Asterand Technologies (Detroit, MI). RNA was extracted from two 10 μm -thick paraffin-embedded tissue sections (Stratagene Absolutely RNA FFPE Kit P/N 400809) utilizing an 18-hour Proteinase K treatment time. RNA was extracted from 30 mg samples of fresh frozen tissue (Stratagene Absolutely RNA Miniprep Kit P/N 400800). The total RNA samples were DNase-treated and evaluated for the presence of contaminating DNA by SYBR Green qPCR analysis. Briefly, 100 ng of each RNA sample was assayed in a Brilliant SYBR Green QPCR Master Mix (Stratagene P/N 600548) using the DNA-specific Quantos QPCR Normalization Primers (Set 1) found in the Stratagene SideStep Kit (Stratagene P/N 400908) and analyzed on the MX3000P Real-Time PCR System (Stratagene P/N 401403). No amplification product was generated during the QPCR reaction, which indicated that the RNA samples were free of genomic DNA contaminants.

RNA amplification and cDNA labeling

Amplified cDNA was generated with the the TransPlex Whole Transcriptome Amplification system (Sigma P/N WTA1-50RXN). An input mass of 50 ng from both the FFPE-derived total RNA and from the fresh frozen-derived total RNA was used in each WTA reaction. In recent experiments, input amounts of 25 ng total RNA from FFPE samples and 5 ng total RNA from fresh frozen samples have been successfully employed. The TITANIUM *Taq* DNA polymerase (Clontech P/N 639208) was used during the "Library Amplification" step in the WTA protocol. Three individual technical replicates were run for each of the four different sample types (FFPE-Normal, FFPE-Tumor, FF-Normal, and FF-Tumor).

Amplified cDNA was purified using the QIAquick PCR Purification Kit (QIAGEN P/N 28106) and the cDNA yields were quantified by NanoDrop spectroscopy, and the quality was assessed using the Agilent 2100 Bioanalyzer. The Agilent Genomic DNA ULS Labeling Kit (Agilent P/N 5190-0419) was used to chemically label 1.5 μg of cDNA with ULS-Cy3 dye for 30 minutes at 85°C. Cy3-labeled samples were purified using Agilent-KREApure columns (a component in the Agilent Genomic DNA ULS Labeling Kit).

Hybridization and washing

The purified, Cy3-labeled 1.5 μg cDNA samples were combined with Agilent 10x Blocking Agent and Agilent 2x Hi-RPM Hybridization Solution (both components of the Agilent Gene Expression Hybridization Kit P/N 5188-5242). Prior to array hybridization, hybridization mixtures were denatured at 95°C for 3 minutes. Agilent-CGHblock (a component of the Genomic DNA ULS Labeling Kit) was added and samples were hybridized to the Agilent Whole Human Genome 4x44K Microarrays (P/N G4112F, AMADID 014850). Each slide contains four identical microarrays consisting of 41,000 unique *in situ* synthesized 60-mer oligonucleotide biological probes. Hybridization was carried out at 20 RPM at a temperature of 65°C for 17 hours before washing in Agilent Gene Expression Wash Buffer 1 (P/N 5188-5325) and Agilent Gene

Expression Wash Buffer 2 (P/N 5188-5326), using the procedures outlined in the Agilent One-Color Microarray-Based Gene Expression Analysis manual (Version 5.7 P/N G4140-90040).

Microarray scanning and data analysis

Scanning and image analysis were performed using the Agilent DNA Microarray Scanner (P/N G2565BA) equipped with extended dynamic range (XDR) software according to the Agilent Gene Expression Users Guide (Version 5.7, P/N G4140-90040). Feature Extraction Software (Version 9.5) was used for data extraction from raw microarray image files. Data visualization and analysis was performed with GeneSpring GX (Version 10.0) software and detected calls were made by employing the default flag settings for Agilent One-Color microarrays.

Results

RNA yield and purity

The RNA yield from the human colon FFPE and fresh frozen tissues were evaluated using a NanoDrop spectrophotometer (**Table 1**) and the RNA quality was assayed using the Agilent 2100 Bioanalyzer (**Figure 1**). The RNA yield obtained from two 10 μm -thick FFPE human colon tissue sections (total surface area $\sim 2\text{ cm}^2$) was 7 to 33 μg as compared to 15 to 25 μg for the 30 mg frozen sample sections. Both the

Storage	Condition	Conc. (ng/ μl)	A_{260}/A_{280}	A_{260}/A_{230}	Elution Vol. (μl)	Yield (μg)
FFPE	Normal	228.59	2.08	1.92	30	6.9
FFPE	Tumor	1098.32	2.06	2.09	30	32.9
FF	Normal	372.55	2.11	2.03	40	14.9
FF	Tumor	634.92	2.13	1.96	40	25.4

Table 1. The Absolutely RNA FFPE and the Absolutely RNA Miniprep Kits produce pure RNA. The A_{260}/A_{280} and A_{260}/A_{230} ratios are >1.90 indicating that pure RNA was isolated with very few contaminants.

FFPE-derived and fresh frozen-derived total RNA samples had high purity ratios ($A_{260/280}$ ratio and $A_{260/230}$ ratio) indicating the total RNA isolated with the Stratagene Absolutely FFPE RNA Extraction Kit was of sufficient quantity and quality for downstream gene expression microarray analysis.

Gene list concordance between FFPE and fresh frozen samples

Lists of the biological genes detected using Agilent's GeneSpring GX 10.0 software were generated for each of the three of the replicate arrays in each of the four storage and disease conditions. The number of genes detected in each condition are indicated under the sample type headings in **Figure 2**. As expected, more genes were detected from the fresh frozen samples than the FFPE samples, with on average 30% more genes detected in the fresh frozen storage condition. Venn diagrams generated with the detected gene lists for each of the disease types illustrate the very high overlap between the FFPE and fresh frozen gene lists. With the normal colon sample, 96.8% of the probes on the FFPE detected gene list were also on the fresh frozen detected gene list. With the tumor colon sample, 95.2% of the probes on the FFPE detected gene list were also on the fresh frozen detected gene list (Figure 2).

Concordance of technical replicate samples

One of the quality control methods available in GeneSpring GX 10.0 and applied here is a principal component analysis (PCA) using *all* of the data (both detected and not detected) from all 12 of the microarrays in this experiment. The replicate tumor samples are clearly distinguished from the replicate normal samples by unsupervised PCA of the FFPE (blue) and fresh frozen (red) samples as shown in **Figure 3**.

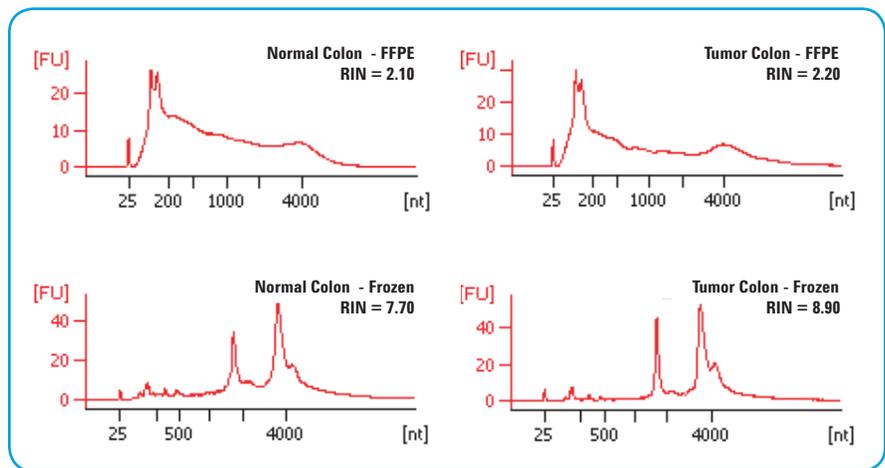


Figure 1. RNA quality was analyzed using the Agilent 2100 Bioanalyzer. The majority of RNA fragments isolated from the FFPE normal colon (NAT) and FFPE colon tumor tissues were between 100–4,000 bp. The low RNA integrity number (RIN, ~2.0) was typical for FFPE extractions as shown in the top two traces.

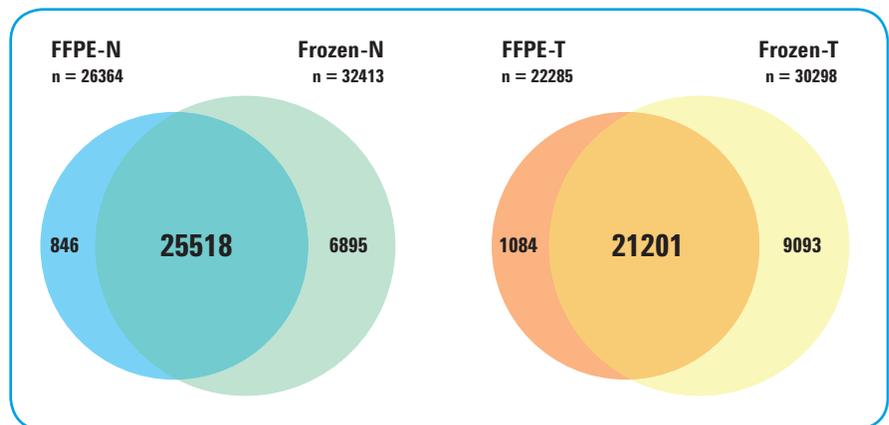


Figure 2. Venn diagram of detected genes in the FFPE and fresh frozen samples for both the normal colon tissue and the colon tumor tissues illustrates high concordance between storage methods. The number of genes detected for each storage condition and disease state is indicated under the titles above the circles in the Venn diagram. The middle number in each overlapping circle indicates concordant calls between the two storage methods.

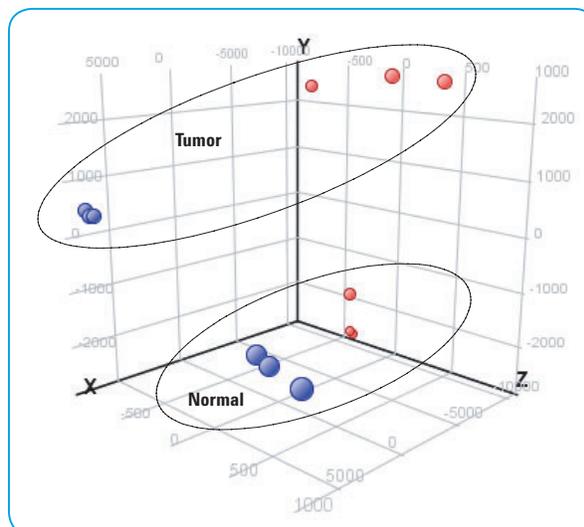
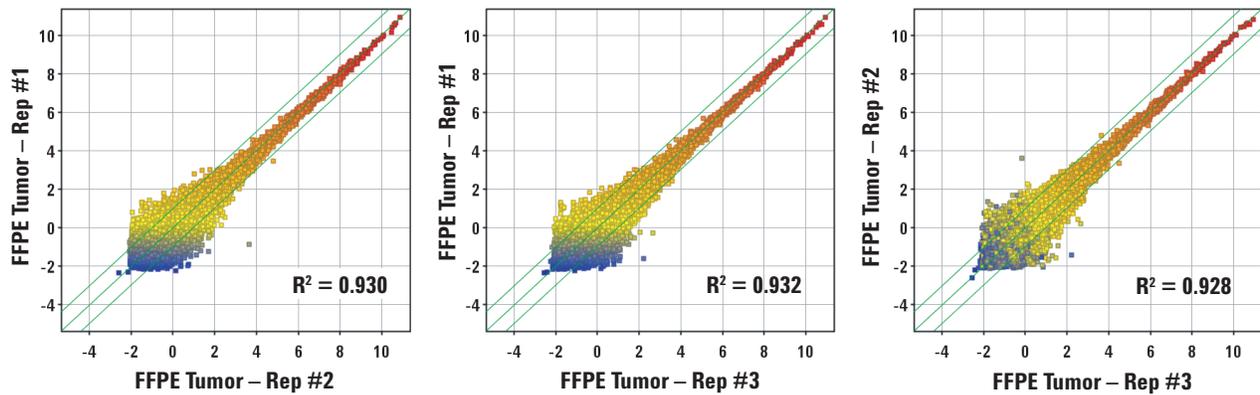


Figure 3. Principal component analysis (PCA) performed in GeneSpring GX 10.0 from the microarray gene expression patterns demonstrate clear disease separation between the disease state of each tissue type. The tight grouping of the technical replicates of the FFPE (blue) and fresh frozen (red) samples also demonstrate good concordance of the replicate microarray assays.

4A: FFPE Correlations



4B: Fresh Frozen Correlations

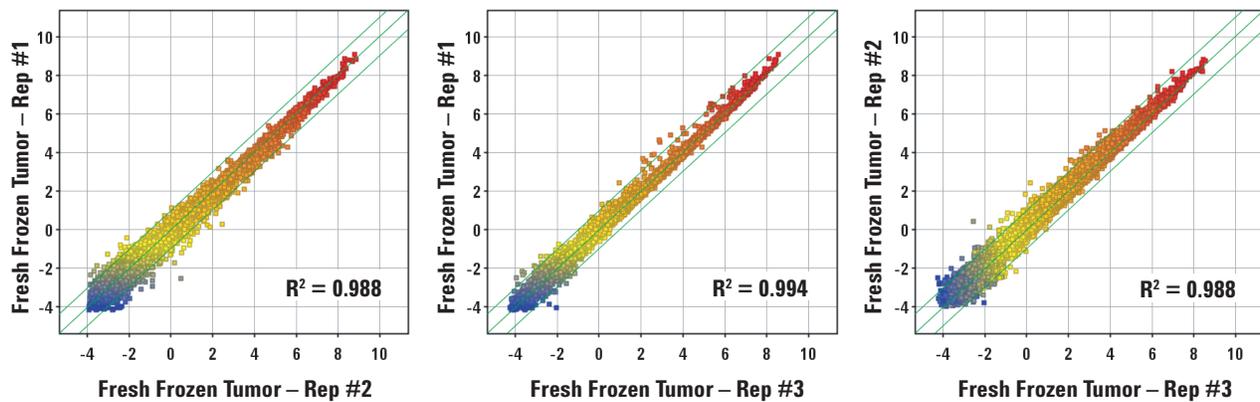


Figure 4. Highly reproducible \log_2 75th percentile-normalized microarray signals were obtained for both the FFPE and the fresh frozen triplicate technical replicate microarrays for the 15,026 genes concordantly detected across all microarrays. Figure 4A shows the technical replicates of the FFPE-stored tumor samples where correlations (R^2) values for the replicate FFPE samples averaged 0.930. Figure 4B shows the technical replicates of the fresh frozen colon tumor samples where R^2 values for the replicate fresh frozen samples averaged 0.990. Results were similar for the normal (NAT) samples (data not shown).

Assay reproducibility can also be demonstrated by comparing the \log_2 of the 75th percentile-normalized microarray signals from technical replicate experiments across the dynamic range of the assay. Both the FFPE assays and the fresh frozen assays span dynamic ranges of four or more orders of magnitude and microarray signals show strong correlations across replicate reactions (Figure 4).

Statistically significant gene expression between normal and tumor samples

An unpaired T-test analysis was performed on the significance of the fold change (tumor versus normal) of the 15,026 genes that were detected on all 12 microarrays. An asymptotic p-value computation was performed with Benjamini-Hochberg FDR multiple testing correction. The resulting data are presented in a volcano plot where the

magnitude of fold change between the normal and tumor conditions are compared to the statistical significance of the fold change (Figure 5). Genes that are significantly expressed (corrected p-value of <0.05) at a two-fold change cutoff are colored red. In both the FFPE and fresh frozen storage conditions, at least 2,200 genes are significantly differentially expressed between the tumor and normal samples.

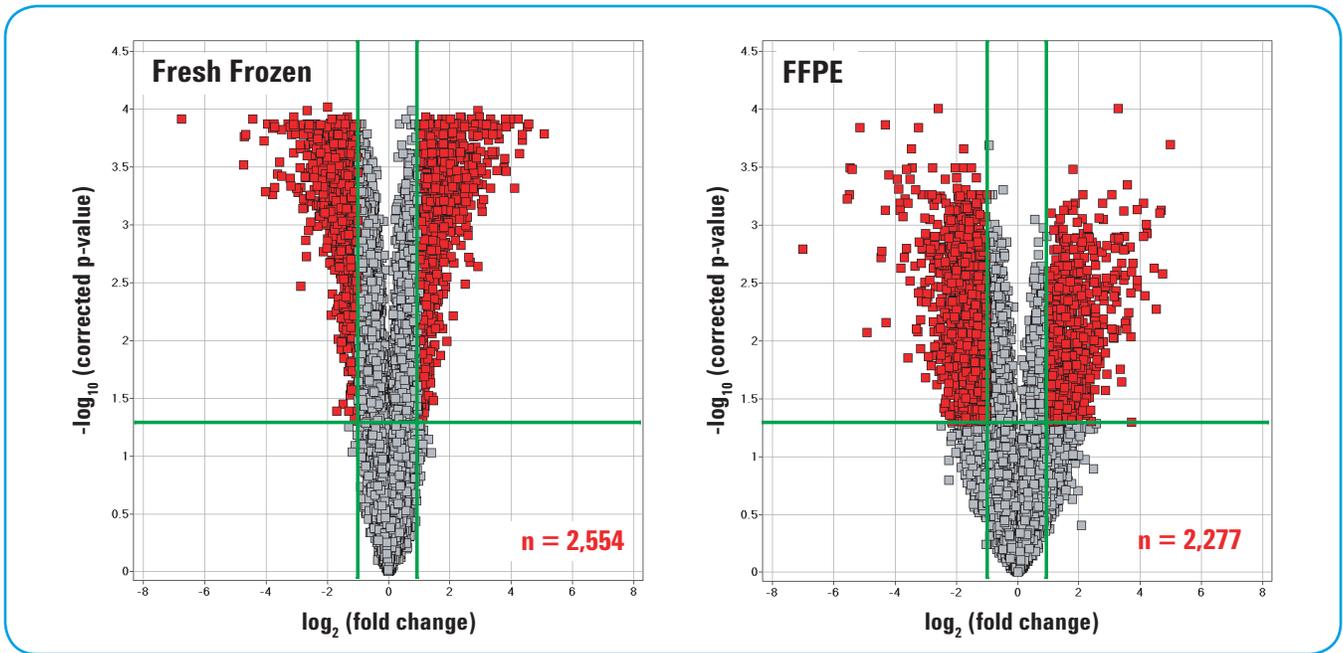


Figure 5. Volcano plot analysis of the differential expression of the fresh frozen and FFPE storage conditions reveals thousands of differentially expressed genes. The \log_2 fold change values are plotted on the x-axis of the volcano plots and are compared to the negative \log_{10} corrected p-values on the y-axis. Genes with an absolute differential expression fold change of at least 2-fold with a corrected p-value of at least 0.05 are colored red.

Concordance of tumor vs. normal ratios between FFPE and fresh frozen samples

Lists of significantly differentially expressed genes for the two tissue storage conditions (red data from the two graphs in **Figure 5**) were compared, and the \log_2 ratios for the 1,141 genes significantly differentially expressed in both the FFPE and fresh frozen samples were plotted (**Figure 6**). Overall, the \log_2 ratios derived from the FFPE samples and the fresh frozen samples are consistent with only 20 genes showing anti-correlation in the direction of fold-change between the two storage conditions.

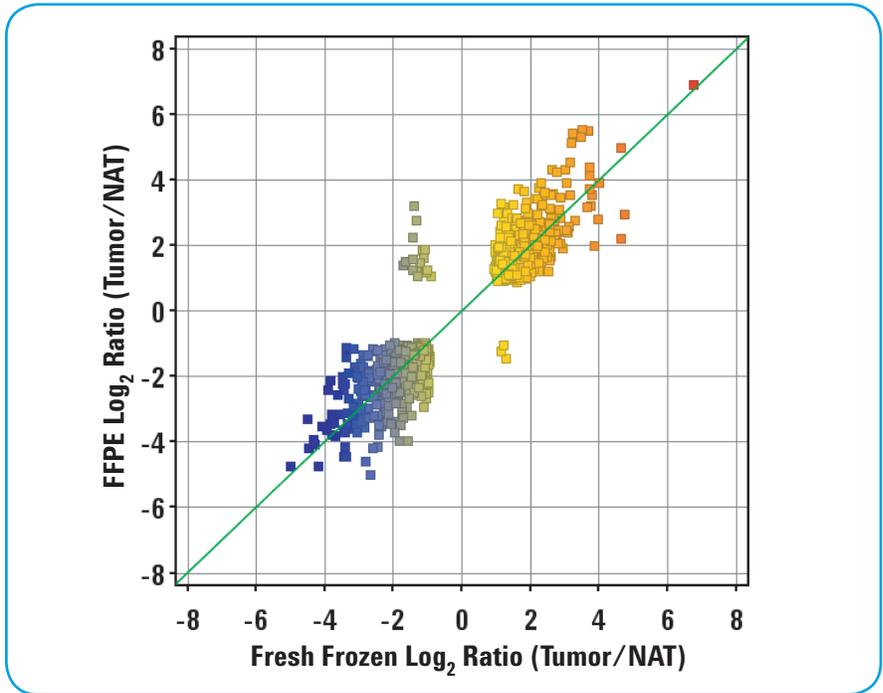


Figure 6. Comparison of the genes significantly differentially expressed (tumor versus normal) in FFPE and fresh frozen samples demonstrates good concordance of expression. The \log_2 ratio values for the fresh frozen samples are plotted on the x-axis and the \log_2 ratio values for the FFPE samples are plotted on the y-axis. A linear regression fit of the data points result in a line with a near-perfect slope of 1.01 and a R^2 correlation of 0.866. Green line represents $x = y$ fit.

Microarray data reproducibility from older FFPE blocks

To assess microarray assay performance with FFPE blocks that were greater than a few years old, we performed replicate assays with a colon tissue sample that was stored in an FFPE block for over 10 years. RNA was isolated (using the same procedure as described) from two 10 μm slices for each of the three replicate isolations. RNA yield from these older samples (average of 2.5 μg per isolation) is reduced compared to the newer samples used the earlier experiment, but is still sufficient for the 25 ng input mass used in the amplification reactions. The replicate isolations were amplified and labeled, and the samples were hybridized to microarrays. The resulting intensity data were compared

in GeneSpring GX 10.0 as described above. More than 50% of the genes were detected concordantly in all three microarrays (21,402/41,000) in this experiment. The \log_2 normalized signals for the 21,402 genes detected across all three microarrays are compared in **Figure 7**. The data demonstrate good reproducibility of microarray signals from the replicate RNA isolations from older FFPE blocks.

Conclusion

The methods described in this application note, as summarized in **Box 1**, enable global gene expression profiling and analysis from degraded RNA from FFPE tissue sources. The data presented in this application note exemplifies the quality of the gene expression infor-

mation that can be generated from FFPE samples and compares it to that obtained from high quality fresh frozen RNA sources. Despite the expected loss of sensitivity when using FFPE RNA, the differential expression results generated with FFPE and fresh frozen samples are highly concordant.

Given the number of archived FFPE tissue samples available for retrospective and prospective studies and the value these samples bring in the ability to link gene expression profiles to clinical outcomes, advances in performing microarray expression profiling on these types of challenging RNA sources are extremely valuable to the research community.

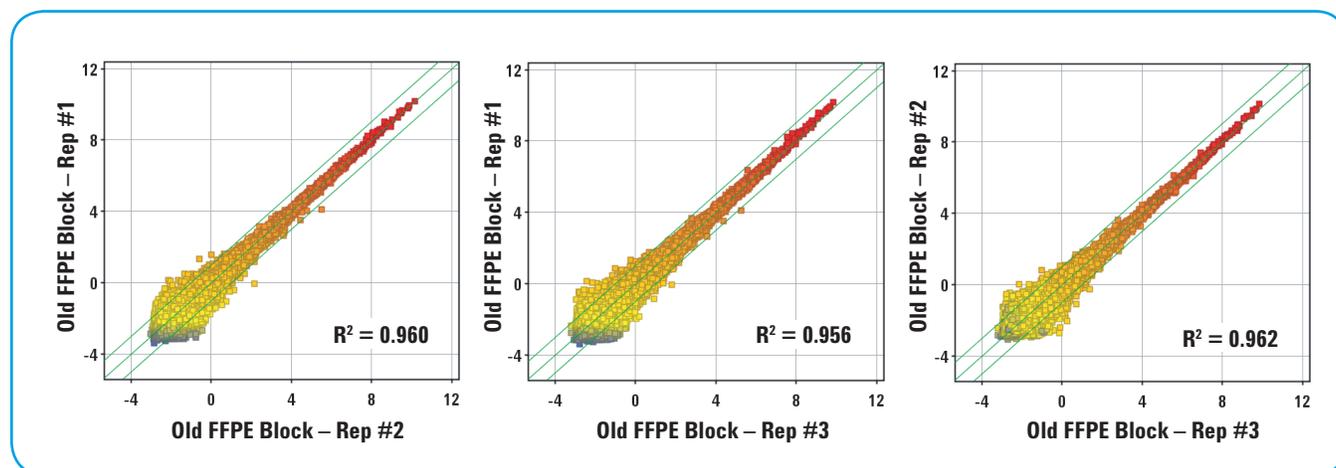


Figure 7. Agilent Gene Expression FFPE microarray workflow demonstrates good reproducibility from older FFPE blocks. The scatter plots demonstrate high correlations (average = 0.960) of \log_2 , 75th percentile-normalized microarray signals for replicate RNA preparations from an 11-year-old FFPE block. In all cases, at least 50% of the genes on the microarray were detected and microarray signals spanned over four orders of magnitude. On average, RNA yield from these older samples ($2.53 \pm 0.1 \mu\text{g}$) was reduced compared to the newer samples.

References

1. van't Veer LJ *et al.*, (2002) **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 415:530-536
2. Glas AM *et al.*, (2006) **Converting a breast cancer microarray signature into a high-throughput diagnostic test.** *BMC Genomics* Oct 30;7:278
3. Paik S *et al.*, (2005) **Technology Insight: application of molecular techniques to formalin-fixed paraffin-embedded tissues from breast cancer.** *Nature Clinical Practice Oncology* (2)5:246-254
4. Masuda N *et al.*, (1999) **Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples.** *Nucleic Acids Research* (27)22:4436-4443

Acknowledgements

The authors wish to thank Petula D'Andrade, Paula Costa, Natalia Novoradovskaya, and Stephanie Fulmer-Smentek of Agilent Technologies for their valuable contributions to this project.

About Agilent Technologies

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms, and speed drug discovery. Engineered for sensitivity, reproducibility, and workflow productivity, Agilent's life science solutions include instrumentation, microfluidics, software, microarrays, consumables, and services for genomics, proteomics, and metabolomics applications.

For more information:

www.opengenomics.com

Find an Agilent customer center in your country:

www.agilent.com/chem/contactus

U.S. and Canada

1-800-227-9770

agilent_inquiries@agilent.com

Asia Pacific

adinquiry_aplsca@agilent.com

Europe

info_agilent@agilent.com

This item is intended for Research Use Only. Not for use in diagnostic procedures. Information, descriptions, and specifications in this publication are subject to change without notice.

Agilent Technologies shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance or use of this material.

© Agilent Technologies, Inc., 2009

Published in USA, April 13, 2009

5990-3917EN



Agilent Technologies